Artificial Neural Network (ANN) For Flood Forecasting At Kasol In The River Satluj, India

Abhinav Sharma[1], Anshu Sharma[2]
[1]Dept of Electronics and Communication Engineering
[2]Dept of PDP
[1, 2] Graphic Era Hill University, Dehradun, Uttarakhand 248001

Abstract- *Flood causes considerable damage to human lives and property almost every year. Therefore, Flood Forecasting & warning has been recognized as the most important, reliable and cost effective non-structural measures for flood mitigation. One of the approaches to flood forecasting is based on modeling the statistical relationship between the hydrologic input and output, without explicitly considering the relationships that exist among the involved physical processes. Example of which is an Artificial Neural Network (ANN), that provides a quick and flexible approach for data integration and model development. In this work, the development of an ANN model for flood forecasting for Satluj river at Kasol station has been performed. The statistical parameters such as auto correlation, partial correlation and cross correlation of the series have been computed and used to select the input vector of the model. The lead-times considered in the model development for flood forecasting are 1 day to 7 days lead. The results of the ANN models are analyzed using statistical indices such as coefficient of correlation, Root Mean Squared Error (RMSE), percentage error in peak flow estimation. The overall results of ANN models indicate that the developed ANN model structures simulate the nonlinearity in the data with reasonable accuracy. There is deterioration in the performance of the models as lead time is increased from 1 to 7 days.*

Keywords- ANN/ Artificial Neural Networks, Flood Forecasting, Back Propagation Algorithm

I. INTRODUCTION

India has also suffered the damage of agricultural crops, transport and communication networks, infrastructure and also to human lives due to floods over the years. An area of more than 40 million in India has been identified as flood prone.[3] India, is a land where a large area is cultivated by various river systems. Most of these rivers are seasonal and are prone to flood. The country experiences severe floods in perennial rivers of northern and middle India during the monsoon season (June to September) due to high intensity of rainfall in the catchment. Therefore there is utmost need to develop real time flood forecasting system to prevent the severe loss due to the natural disaster of flood. There can be some conventional rainfall- runoff models developed to target the problem. But development and calibration of these requires to optimize a number of physical parameters interacting in a complex way and therefore a large amount of historic data is needed. Instead, black box models are proven to produce good result from the input-output mapping when a detailed physical description of the process is not required. In recent years ANN models have been applied successfully for flood forecasting because of its ability to map any nonlinear function of given sufficient complexity. ANNs are proven to produce improved performance over other black box models in numerous hydrological studies (Hsu et al., 1995). The ANN models are advantageous as for development of these models the mathematical form of the information about the complex interaction among the physical variables is not required.

II. THE STUDY AREA

Study Area and Data Availability

The catchment area of the Sutlej river up to Kasol was considered for this study. The study area is located on the upstream of Bhakra reservoir. The catchment area up to Kasol is 56980 sq.km. The location of the reservoir is presented in Fig. 1. For the current application, the daily rainfall values for Kalpa, Rampur, Rackchham, Berthin, Bhakra, Kahu, Kasol, Kaza, Namagia, and Suni were available from the year 1987 to 2000. The rainfall values of all the stations have been used. The maximum rainfall values recorded at Kalpa, Rampur, Rackchham, Berthin, Bhakra, Kahu, Kasol, Kaza, Namagia, and Suni during 1987 to 2000 were 104, 86, 66, 213.2, 281.5, 224, 141, 80, 42.8 and 170 mm respectively.

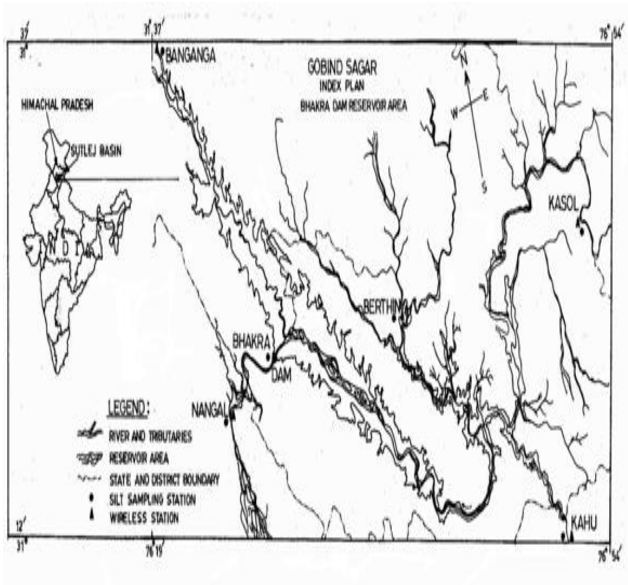

Fig. 1. Index map of Bhakra Reservoir

The rainfall stations Berthin, Bhakra and Kahu received heavy rainfall in a single day than other rainfall stations and are located nearer to the reservoir. The rainfall stations Rampur, Suni and Kasol are located along the course of the Sutlej river. The discharge values at Kasol for the same period were also available and the discharge measured on 01 August 2000 was 2689.28 cumecs. From correlation analysis it was learnt that the rainfall values at Kasol, Rampur and Suni and discharge at Kasol had significant relationship on streamflow at Kasol.

III. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks are the biologically inspired simulations performed on the computer to perform certain specific tasks like clustering, classification, pattern recognition etc.. Artificial neural networks can be viewed as weighted directed graphs in which artificial neurons are nodes and directed edges with weights are connections between neuron outputs and neuron inputs and are discussed in detail in a number of hydrologic papers. For example, Kishore and Kaur, 2012, Thirumalaiah, K., and Deo,1998, Prachatos Mitra et al, 2016, Sulafa HagElsafi, 2014 Portugal, 1995; Minns and Hall, 1996; See et al, 1997; Danh et al, 1998.The Artificial Neural Network receives input from the external world in the form of pattern and image in vector form. These inputs are mathematically designated by the notation x(n) for n number of inputs. Each input is multiplied by its corresponding weights. Weights are the information used by the neural network to solve a problem. Typically weight represents the strength of the interconnection between neurons inside the neural network. The weighted inputs are all summed up inside computing unit (artificial neuron). In case the weighted sum is zero, bias is added to make the output not- zero or to scale up the system response. Bias has the weight and input always equal to '1'.The sum corresponds to any numerical value ranging from 0 to infinity. In order to limit the response to arrive at desired value, the threshold value is set up. For this, the sum is passed through activation function. The activation function is set of the transfer function used to get desired output. There are linear as well as the non-linear activation functions.

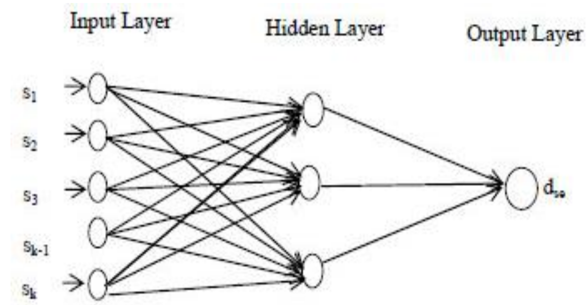

Fig.2-Three-layered feed forward ANN

The most common type of artificial neural network consists of three groups, or layers, of units: (1) a layer of "input" units is connected to (2) a layer of "hidden" units, which is connected to (3) a layer of **"**output" units. A typical three-layered feed forward ANN is shown in Fig. 2

The activity of the input units represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input units and the hidden units. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden units and output units.

III-I- Modeling a Neuron

A neuron is connected to other neurons via its input and output links. Each incoming neuron has an activation value and each connection has a weight associated with it. The neuron sums the incoming weighted values, representing the input from the neuron.

In order to train a neural network, the weights of each unit are adjusted in such a way that the error between the desired output and the actual output is reduced. This process requires that the neural network compute the error derivative of the weights (ED). In other words, it must calculate how the error changes as each weight is increased or decreased slightly. The back propagation algorithm is the most widely used method for determining the error derivative of the weights (Zweiri, Y. H. et al 2003, Xinghuo Yu et al 2002)

Back-propagation (BP) is a well known network that has been known for its accuracy because it allows itself to learn and improving itself thus it can achieve higher accuracy (Yeremia, Hendy, et al, 2013).The algorithm computes each ED by first computing the EAL, the rate at which the error changes as the activity level of a unit is changed. For output units, the EAL is simply the difference between the actual and the desired output. To compute the EAL for a hidden unit in the layer just before the output layer, we first identify all the weights between that hidden unit and the output units to which it is connected. We then multiply those weights by the EAs of those output units and add the products. This sum equals the EAL for the chosen hidden unit. After calculating all the EALs in the hidden layer just before the output layer, we can compute in like fashion the EALs for other layers, moving from layer to layer in a direction opposite to the way activities propagate through the network. This is what gives back propagation its name. Once the EAL has been computed for a unit, it is straight forward to compute the ED for each incoming connection of the unit. The ED is the product of the EA and the activity through the incoming connection. The literature by Rumelhart et al, 1986 can be referred for the details of the algorithm.

IV. DATA ANALYSIS AND MODEL SET-UP

For the development of model, daily rainfall data for Kahu, Kasol, Rampur, Suni, Berthin, and Bhakhara was available for the period from 01-01-1987 to 31-12-2004. The daily discharge data for Kasol was also available for the same period. These rain gauge and river gauge stations are marked in Fig 1. In this study the ANN models have been developed to forecast the river flow at Kasol at 1 day to 7 Day in advance.

To set-up the ANN model the input vectors are taken as rainfall and discharge values of stations which are to the upstream of the target station. By finding the lags of rainfall and discharge values that significantly influence the forecasted flow, the number of antecedent rainfall and discharge values are determined. These values that correspond to different lags can be determined accurately by performing statistical analysis of the data series. The input vector is elected generally by trial and error method; however, Sudheer et al. (2002) have presented a statistical procedure that avoids the trial and error procedure. They reported that the statistical parameters such as auto correlation function (ACF), partial auto correlation function (PACF) and cross correlation function (CCF) can be used for this purpose. It is evident from Fig. 3, which presents the ACF plot of river flow at Kasol, that it is autoregressive. The PACF of flow series at Kasol (Fig 4) gives potential antecedent runoff values that have influence on the runoff value at the current period. It can be seen from the Fig 4 that the runoff series up to 2 lags should be included in the input vector. The PACF of the flow series at Kasol and CCF of flow series at Kasol between rainfall series at Rampur, Suni, Berthin, Kahu, Kasol and Bhakhara suggest the input vector to the ANN model. From the Fig. 5 the cross correlation between the spatially averaged rainfall and runoff at Kasol indicates that the rainfall at1 lag influences the runoff. In the same way, the figures 6, 7, 8 and 9 present the influencing lags of runoff series at Rampur, Suni, Berthin, Kahu and Bhakhara. On the basis of PACF and CCF of the data series, the following input vector is selected for neural network training-

$$Q_{Kasol,t} = f\,(Q_{Kasol,t-1},\ Q_{Kasol,t-2},\ R_{Rampur,t},\ R_{Berthin,t},\ R_{Bhakhara,t},\ R_{Kahu,t},\ R_{Kasol,t},\ R_{Suni,t}) \qquad 1$$

In which Q and R are discharge and spatially averaged rainfall values respectively.

Evaluation of Performance

For performance evaluation of ANN model the whole data length is divided into two, one for calibration (training) and another for validation of artificial neural network model. The performance during calibration and validation is evaluated by performance indices such as root mean square error (RMSE), model efficiency (Nash and Sutcliffe, 1970) and coefficient of correlation (R). They are defined as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{K}(t-y)^2}{K}} \qquad 2$$

$$\text{Efficiency} = 1-\frac{\sum (t-y)^2}{\sum (t-\bar{t})^2} \qquad 3$$

$$\text{Coefficient of Correlation} = \frac{\sum TY}{\sqrt{\sum T^2 \sum Y^2}} \qquad 4$$

where K is the number of observations; t is the observed data; y is computed data; $T = t - \bar{t}$ in which $\bar{t}$ is the mean of the observed data; and $Y = y - \bar{y}$ in which $\bar{y}$ is the mean of the computed data.

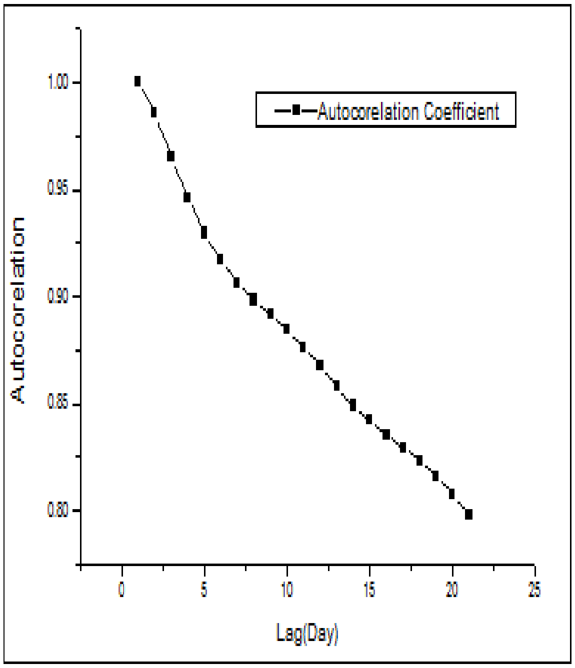

Figure 3 The autocorrelation of the Discharge series at Kasol

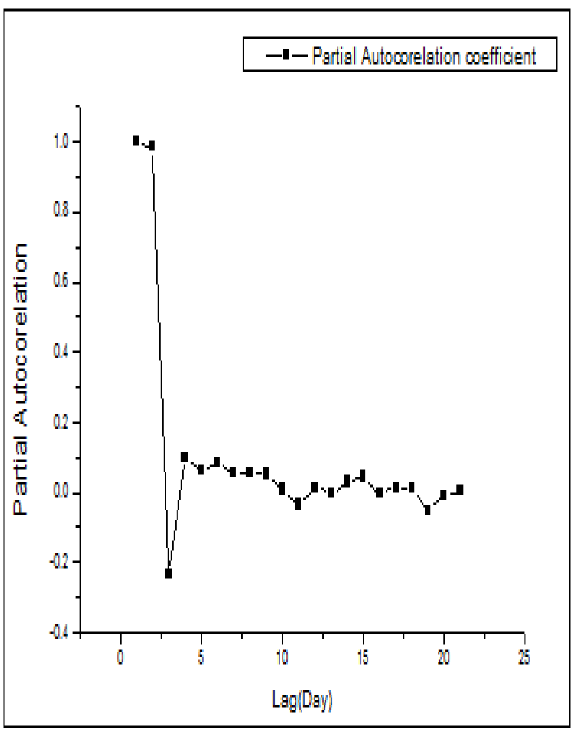

Figure 4 The partial autocorrelation of the discharge series at Kasol

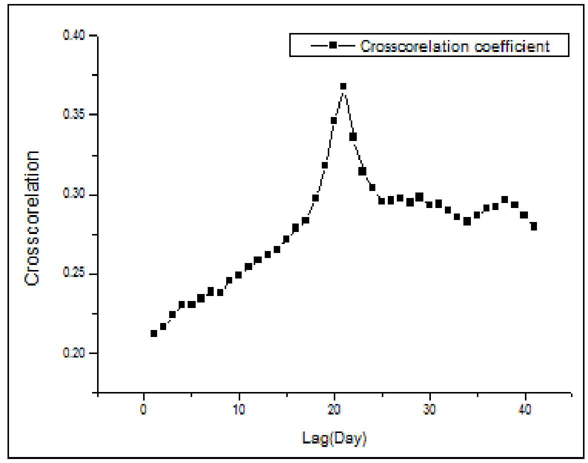

Figure 5 The cross correlation of rainfall at Kasol with discharge series at Kasol

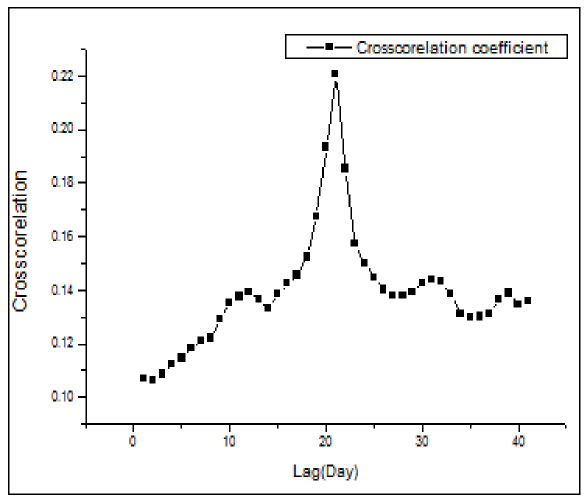

Figure 6 The cross correlation of rainfall at Rampur with discharge series at Kasol

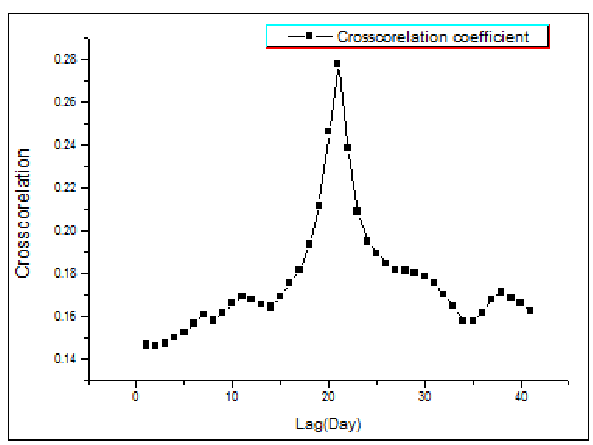

Figure 7 The cross correlation of rainfall at Suni with discharge series at Kasol

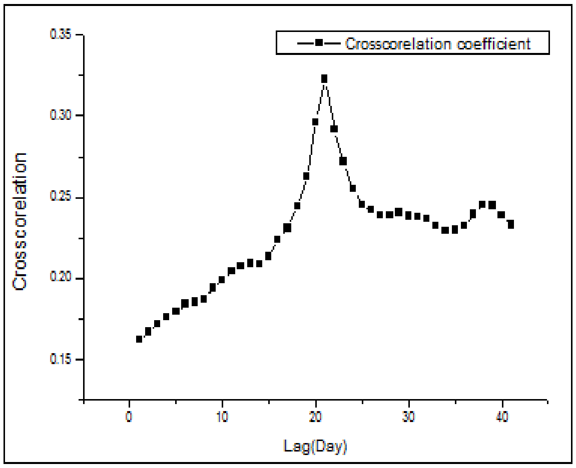

Figure 8 The cross correlation of rainfall at Berthin with discharge series at Kasol

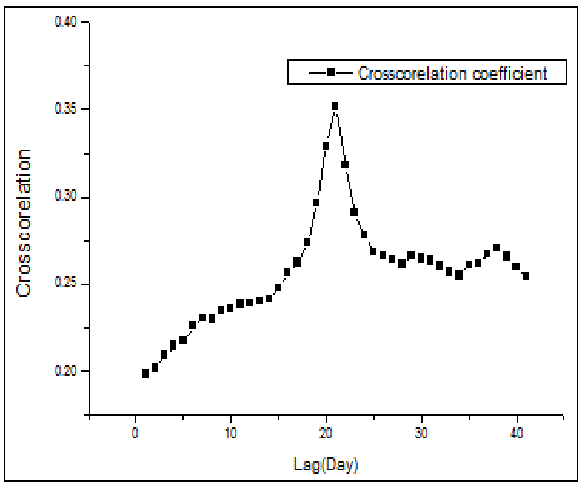

Figure 9 The cross correlation of rainfall at Kahu with discharge series at Kasol

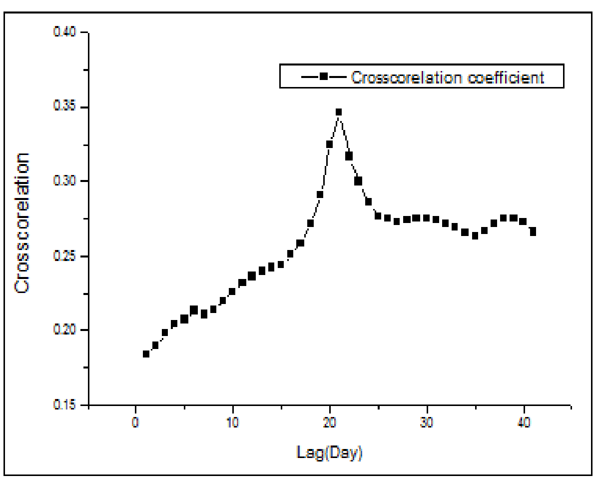

Figure 10 The cross correlation of rainfall at Bhakhara with discharge series at Kasol

Model Set-Up

Back propagation algorithm has been used for setting up the model. 1 day lead to 7 day lead is taken as the lead times for development of models. The data set from the year 1987 to year 2000 is considered for calibration and that from year 2001 to year 2004 is used for validation of the model. MATLAB (The Mathworks, Inc., 2001) has been used as the software for training the models.

V. RESULTS AND DISCUSSIONS

The performances of the forecasts have been summarized in the form of tables. Calibration results and Validation results forecasted have been presented in the form of hydrographs for 1 hour lead-time in Figures 10 and 11 respectively along with the corresponding observed flow series. The viability of the developed ANN model in forecasting the flow at Kasol is clearly depicted in the hydrographs The results are further analyzed using statistical indices too. Table 1, 2, 3 and 4 present the results of the calibration and validation of the ANN models for all the lead-times in terms of various statistical indices.

Table 1 Results of ANN model for 1 day lead time

	Cal	Val
Coefficient of correlation	0.9925	0.9908
RMSE (cumes)	49.56	44.55
Model efficiency	0.9852	0.9815

Table 2 Results of ANN model for 2 day lead time

	Cal	Val
Coefficient of correlation	0.9775	0.9755
RMSE (cumes)	85.94	73.37
Model efficiency	0.9555	0.9498

Table 3 Results of ANN model for 3 day lead time

	Cal	Val
Coefficient of correlation	0.9607	0.9624
RMSE (cumes)	113.06	91.78
Model efficiency	0.9230	0.9215

Table 4 Results of ANN model for 4 day lead time

	Cal	Val
Coefficient of correlation	0.9469	0.9506
RMSE (cumes)	130.91	104.98
Model efficiency	0.8967	0.8974

Table 5 Results of ANN model for 5 day lead time

	Cal	Val
Coefficient of correlation	0.9364	0.9407
RMSE (cumes)	142.98	114.67
Model efficiency	0.8768	0.8776

Table 6 Results of ANN model for 6 day lead time

	Cal	Val
Coefficient of correlation	0.9264	0.9316
RMSE (cumes)	153.42	122.74
Model efficiency	0.8582	0.8598

Table 7 Results of ANN model for 7 day lead time

	Cal	Val
Coefficient of correlation	0.9194	0.9239
RMSE (cumecs)	160.17	128.87
Model efficiency	0.8454	0.8456

As the tables suggest that the highest coefficient of correlation for both calibration and validation has been achieved for table-1 (1 day lead time) it is indicated that developed ANN model is remarkable in estimating the forecasts with reduced error. As the lead time increases from 1 day to 7 days the performance of ANN model deteriorates. The ANN model for 1 day lead-time can be used for the computation of flood forecast at Kasol.

VI. SUMMARY AND CONCLUSION

In this work, the development of an ANN model for flood forecasting for Satluj river at Kasol station was performed using the average daily rainfall data for Kahu, Kasol, Rampur, Suni, Berthin, and Bhakhara and daily discharge data at Kasol from 1987 to 2004. The input vectors of the model were selected using the statistical parameters such as auto correlation, partial correlation and cross correlation of the series. The lead-times considered in the model development for flood forecasting are 1 day to 7 days lead. The results of the ANN models are analyzed using statistical indices such as coefficient of correlation, Root Mean Squared Error (RMSE), percentage error in peak flow estimation. It is concluded from the validation and calibration results of the models that the ANN model for 1 day lead-time can be used for issuing flood warnings at Kasol. There is deterioration in the performance of the models as lead time is increased from 1 to 7 days.

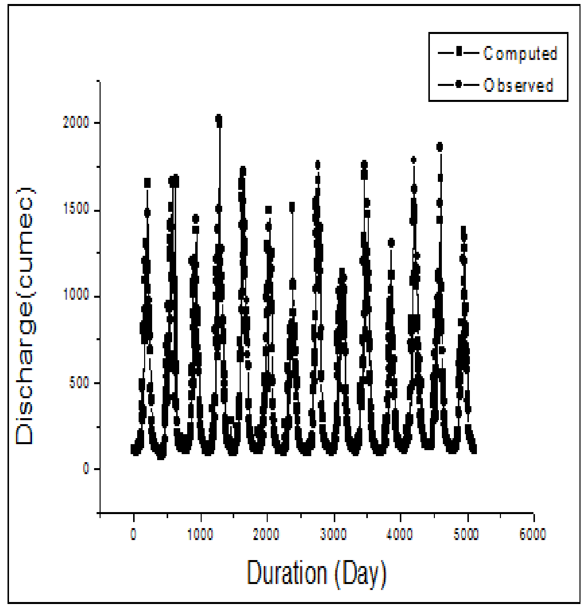

Figure 10 Calibration result of data (1 day lead)

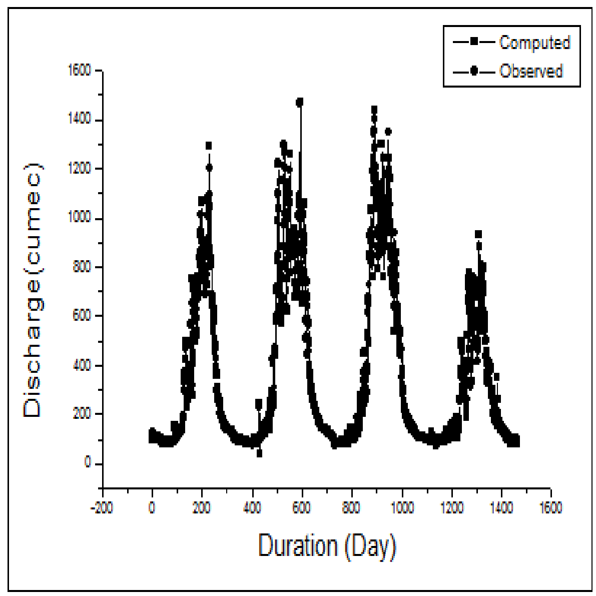

Figure 11 Validation result of data (1 day lead)

REFERENCES

[1] Sulafa Hag Elsafi, “Artificial Neural Networks (ANNs) for flood forecasting at Dongola Station in the River Nile, Sudan”, *Alexandria Engineering Journal*, Volume 53, Issue 3, September 2014, Pages 655-662

[2] Prachatos Mitra,Ronit Ray, Retabrata Chatterjee, Rajarshi Basu,Parmarth Saha, “Flood Forecasting using Internet of Things and Artificial Neural Networks”, Electronics and Mobile Communication Conference (IEMCON), 2016 IEEE , 10.1109/IEMCON.2016.7746363

[3] R.D.Singh, International Hydrology Series, "Hydrological Modeling in Arid and Semi-Arid Areas", edited by Howard Wheater, Soroosh Sorooshian and K.D. Sharma

[4] Zweiri Y.H., Whidborne, J. F. and Seneviratne, L. D. (2003), "A three-term backpropagation algorithm", *Neurocomputing*, 50, 305 - 318. ISSN (print) 0925-2312

[5] Xinghuo Yu, M. Onder Efe, and Okyay Kaynak "A General Backpropagation Algorithm for Feedforward Neural Networks Learning", IEEE Transactions on Neural Networks, vol. 13, no. 1, January 2002.

[6] Kuldip Vora and Shruti Yagnik, "A Survey on Backpropagation Algorithms for Feedforward Neural Networks", *International Journal of Engineering Development and Research* ISSN: 2321-9939

[7] Yeremia, Hendy, et al. "Genetic Algorithm and Neural Network for Optical Character Recognition." *Journal of Computer Science* 9.11 (2013): 1435.

[8] Anderson, Dave, McNeil, George. Artificial neural networks technology. Data and analysis Centre for Software, Rome, August 1992 <http://www.dtic.mil>.

[9] Sajikumar, S., and Thandaveswara, B.S. (1999). "A non-linear rainfall-runoff model using an artificial neural network." *Journal of Hydrology*, 216, 32-55.

[10] See, L., Dougherty, M., and Openshaw, S. (1997). "Some initial experiments with Neural Network Models of flood forecasting on the river Ouse.", *Proceedings of Gecomputation'97 & SIRC'97*, University of Otago, New Zealand, 15-22.

[11] Sudheer, K. P., Gosain, A. K., and Ramasastri, K. S. (2002). "A data-driven algorithm for constructing artificial neural network rainfall-runoff models.", *Hydrological Processes*, Vol. 16, 1325-1330.

[12] Sudheer, K. P., Nayak, P. C., and Ramasastri, K. S. (2003). "Improving Peak Flow Estimates in Artificial Neural Network River Flow Models. *Hydrological Processes*, 17(1), 677-686.

[13] Dr. Rama Kishore,Taranjit Kaur "Backpropagation Algorithm:An Artificial Neural Network Approach for Pattern Recognition", International *Journal of Scientific & Engineering Research*, Volume 3, Issue 6, June-2012 ISSN 2229-5518

[14] The MathWorks, Inc. (2001). "ANN Toolbox User's Guide." 3 Apple Hill Drive, Natick, MA 01760-2098.

[15] Thirumalaiah, K., and Deo, M.C. (1998). "River stage forecasting using artificial neural networks." *Journal of Hydrologic Engineering*, ASCE, 3 (1), 26-32.

[16] Thirumalaiah, K., and Deo, M.C. (2000). "Hydrological forecasting using neural networks." *Journal of Hydrologic Engineering*, ASCE, 5 (2), 180-189.

[17] Tokar, A.S.. and Johnson, A. (1999). "Rainfall-runoff modeling using artificial neural networks". *Journal of Hydrologic Engineering*, ASCE, 4(3), 232-239.

[18] Zealand, C.M., Burn, D.H., and Simonovic, S.P. (1999). "Short term streamflow forecasting using artificial neural networks." *Journal of Hydrology*, 214, 32-48.

[19] Burian, S.J., Durrans, S.R., Nix, S.J. and Pitt, R.E. (2001). "Training artificial neural networks to perform rainfall disaggregation." *Journal of Hydrologic Engineering*, ASCE, 6(1), 43-50.

[20] Elshorbagy, A., Simonovic, S.P. and Panu, U.S. (2000). "Performance evaluation of artificial neural networks for runoff prediction." *Journal of Hydrologic Engineering*, 5(4), 424-433.